Crop Yield Prediction Based on Satellite Images

Blue Marble an impact InsurTech with a mission to bring insurance to the underserved

Crop Yield Prediction Based on Satellite Images

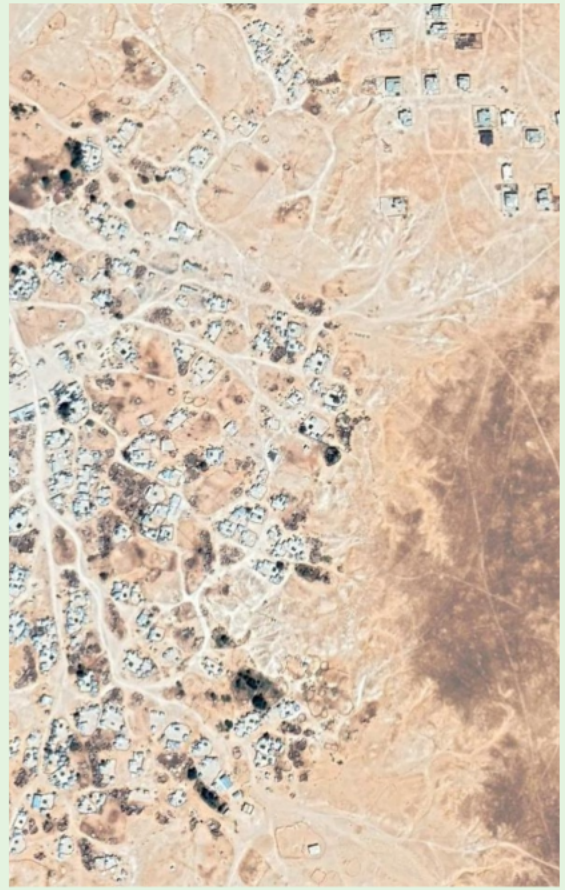
Blue Marble an impact InsurTech with a mission to bring insurance to the underserved

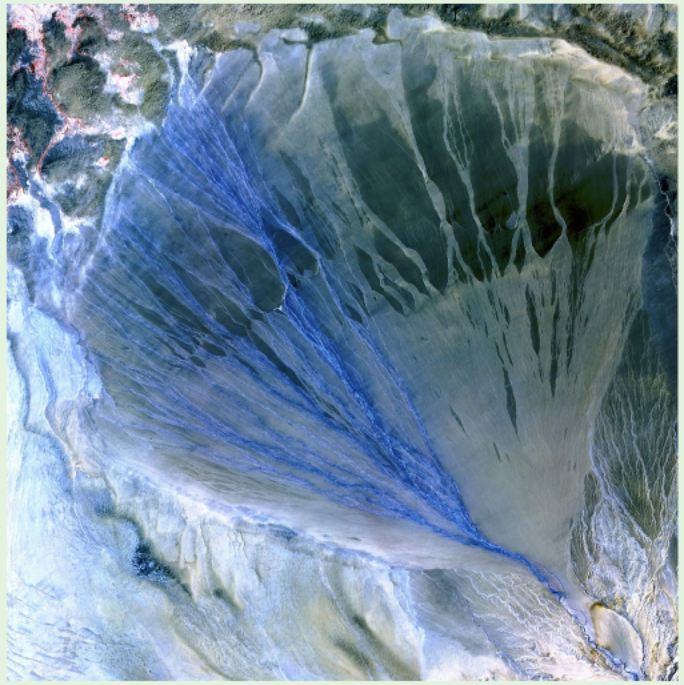
Food Demand Projection and Hunger Risk (2010-2050)

It was estimated that global food demand will increase between 35% and 56% between 2010 and 2050.

At the same time population at risk of hunger will either decrease by 91% or increase by 30%

Agriculture is key to food security, especially in the less developed countreis





Parametric Insurance

There are multiple risk factors that affect the crop yield and that are difficult to manage on the individual level and hence create an opportunity for companies, like Blue Marble, to support farming comuniteis and help redistribute risk on the national and global level.

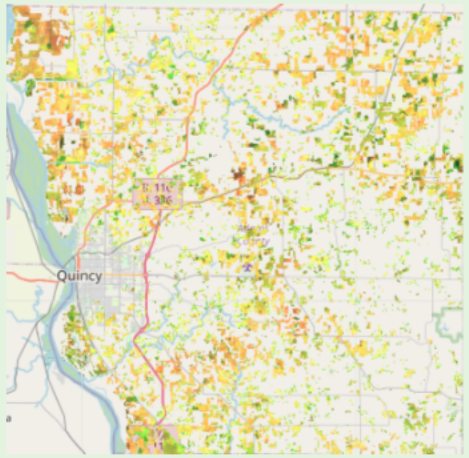
Parametric insurance operates by triggering payouts based on predetermined parameters, such as satellite-based crop yield estimates. Instead of traditional claims processes, farmers receive immediate financial assistance when specific conditions are met, streamlining recovery efforts after adverse events.

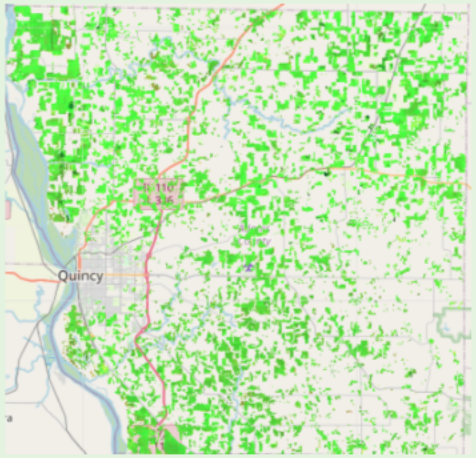
Data collection

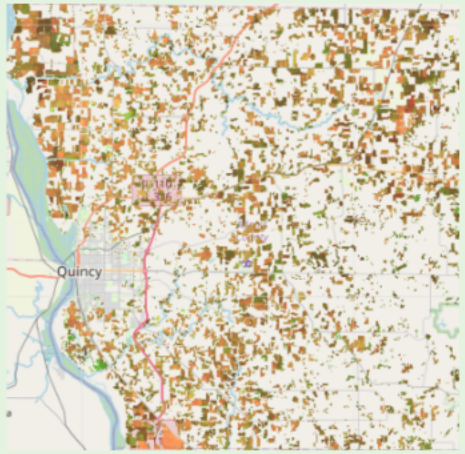
County level corn crop yield in the U.S. for yeras 2016-2022

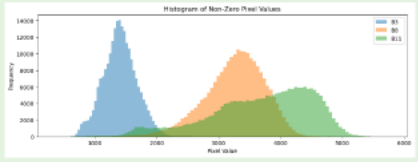
Histograms based on ca. 29,000 Sentinel-2 satellite images providing insights into agriculture through high-resolution imaging.

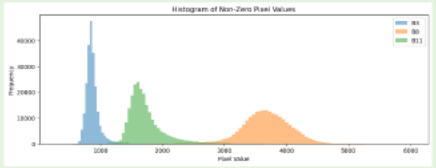
Satellite Image Examples

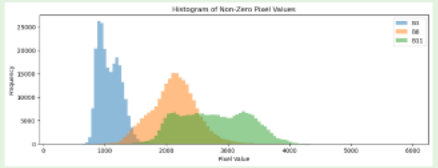












Three satellite images exemplify the different stages of crop growth captured by satellite. These images illustrate how spectral data correlates with season and potential yield outcomes, serving as a foundation for predictive modeling.

Results and Challenges

Results

1 LSTM

RMSE: 24 Correlation: 0.777

2 CNN

RMSE: 26.6 Correlation: 0.759

3 MLP

RMSE: 32.64 Correlation: 0.75

4 Ridge Regression

RMSE: 39.48 Correlation: 0.513

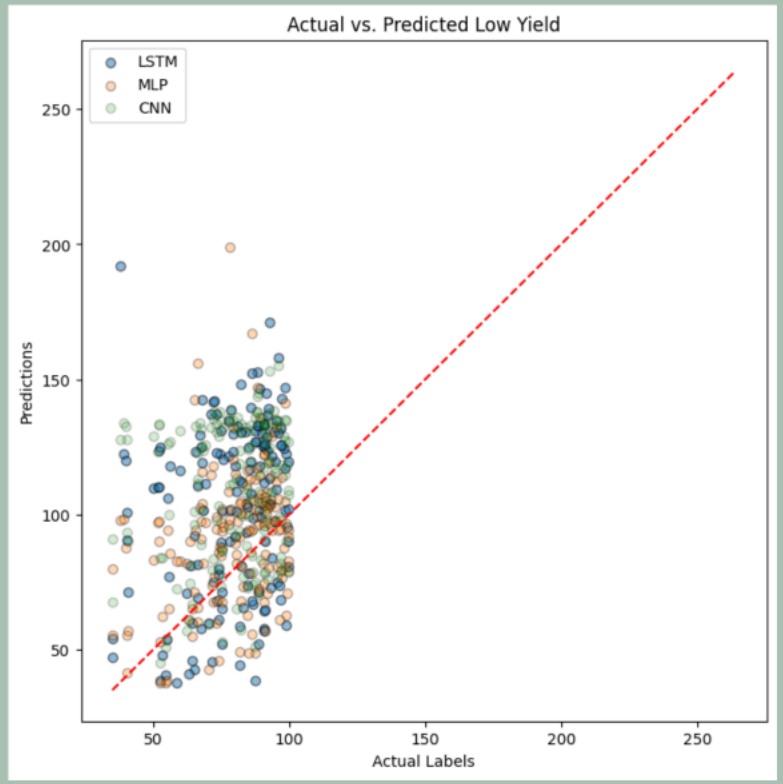
Modeling Approaches

LSTM, CNN, MLP, and Ridge Regression have been applied to crop yield prediction.
LSTM because of architecture designed to capture temporal information. CNN for its efficiency with images. Both have been havily used in scientific papers. MLP and Regressoin as a bseline

Challenges: fit when crop yield is low

Sattelite images carry a lot of useful information.

Performance offered by models, especially for low crop yields, is not good enough to be a foundation of an insurance product. The best performing model had accuracy of 50%.



Data Availability Issues

A major challenge in crop yield prediction is the scarcity of publicly available, accurate, fine-granularity labeled data. Data is easily available for a few differet crops in the U.S. But this is not where parametric insurance could bring the most benefits.

Next Steps

Fixing data issues is very costly and difficult but we can still explor ways to improve models

Improvements to data preprocessing

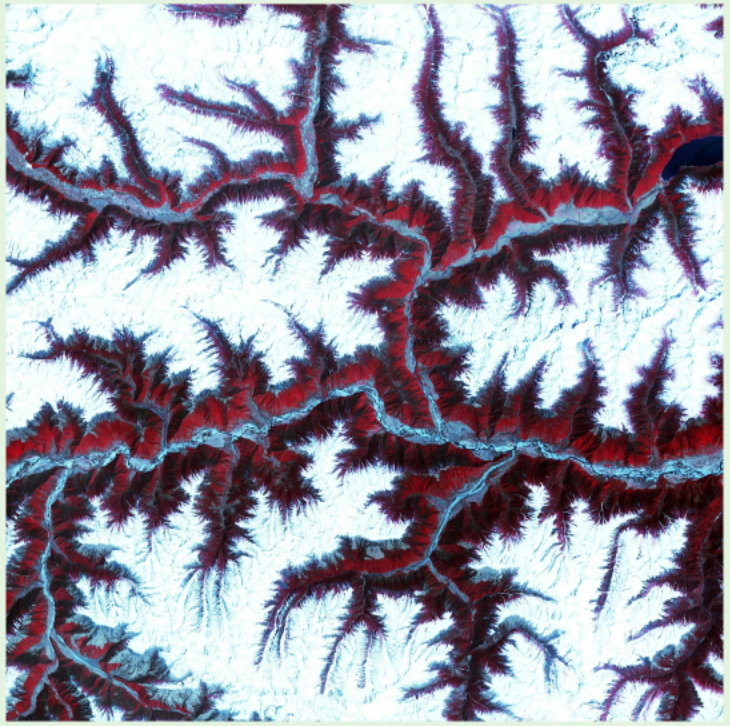
Obtain higher resolution of sattelite images. Currently 60x60m but could be as low as 10x10m ie. 36 times more precise.

Implementing data augmentation techniques, ie create extra synthetic data based on the already collected data.

Exploring other architectures and multiple modalities

Integrating different data modalities, such as combining satellite images with weather data, could offer a more comprehensive picture.

Emplying more compute would allow experimenation with more recent model architectures like transformers.



Questions ?